

Improving the estimation accuracy of titration-based asphaltene precipitation through power-law committee machine (PLCM) model with alternating conditional expectation (ACE) and support vector regression (SVR) elements

Amin Gholami¹ · Omid Mohammadzadeh² · Shahin Kord³ · Siyamak Moradi¹ · Bahram Dabir⁴

Received: 27 February 2015 / Accepted: 2 June 2015 / Published online: 25 June 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Asphaltene precipitation/deposition have been longstanding issues in petroleum industry which lead to decline in oil production and economical efficiency. Owing to severe undesirable issues associated with this phenomenon, it is crucial to develop a reliable, accurate, and robust approach for quantitative estimation of asphaltene precipitation. In the first section of this paper, amount of asphaltene precipitation from stock tank oil through titration process was estimated using two predictive methods of Support Vector Regression (SVR) as well as Alternating Conditional Expectation (ACE). A novel predictive method, the so-called Power-Law Committee Machine (PLCM) with constituents of SVR and ACE, was then employed for estimation of the amount of asphaltene precipitation. PLCM model assigns weight factors to each individual sub-model of SVR and ACE to specify the contribution of each particular model in the overall prediction of asphaltene precipitation. Optimal values of these weight factors were extracted by means of Genetic Algorithm (GA) since it was already inserted as the combiner in the structure of the PLCM model. To validate this predictive tool, experimental data collected from open source literature were compared against the model predictions. It

was observed that PLCM model can estimate the amount of asphaltene precipitation with very high accuracy and it had more satisfactory prediction performance compared to the other models of SVR and ACE.

Keywords Asphaltene Precipitation · Support Vector Regression · Alternating Conditional Expectation · Power-Law Committee Machine

Abbreviations

ACE	Alternating conditional expectation
SVR	Support vector regression
PLCM	Power-law committee machine
GA	Genetic algorithm
EOR	Enhanced oil recovery
ANN	Artificial neural network
SARA	Saturates, aromatics, resins, and asphaltenes
X	Independent variable
Y	Dependent variable
e^*	Error which is not captured by ACE
$f(x, w)$	Decision function
W	Weight factor
B	Bias term
$R(C)$	Risk function
C	Penalty parameter
RBF	Radial basis function
G	Kernel function
CM	Committee machine
R^2	Correlation coefficient
MSE	Mean square error
R_v	Dilution ratio (mL/g)
M_w	Molecular weight of solvent (g/mol)
T	Temperature (K)
W_t	Amount of asphaltene precipitation (wt%)

✉ Omid Mohammadzadeh
Omohamma@uwaterloo.ca

¹ Department of Petroleum Engineering, Petroleum University of Technology, Abadan, Iran

² DBR Technology Center, Schlumberger, Edmonton, AB, Canada

³ Department of Petroleum Engineering, Montan University of Leoben, Leoben, Austria

⁴ Department of Chemical Engineering, Amirkabir University of Technology, Tehran, Iran

β	Regression coefficient
ϕ	Transformation of independent variable
θ	Transformation of response variable
ϕ^*	Optimal transformation of independent variable
θ^*	Optimal transformation of response variable
ε^2	Error variance
ζ, ζ^*	Slack variable
α, α^*	Lagrangian multipliers

Introduction

Crude oil is a complex mixture which makes it very difficult to achieve thorough characterization at the level of individual molecular type (Ahmadi and Shadizadeh 2012). In order to rectify this issue, group type analysis is deemed as a viable option for crude oil characterization. SARA separation technique is an example of such a group type analysis in which the crude oil is characterized in terms of four distinct classes of saturates, aromatics, resins, and asphaltenes. This separation is implemented based on differences in polarity and solubility (Rasuli Nokandeh et al. 2012; Mohammadi et al. 2012). When dealing with crude oil fractions, most of the attention is focused on asphaltenes mainly due to its adverse influences on the downstream and upstream sectors of petroleum industry. Asphaltenes are known as the heaviest and the most polar portion of crude oil (Kurup et al. 2012) and are defined as fractions of crude oil that are soluble in benzene but insoluble in n-heptane (Shirani et al. 2012a, b). Under initial reservoir conditions, asphaltenes are equilibrated in crude oil through peptizing by resins (Kurup et al. 2012). With changes in thermodynamic conditions such as pressure, temperature, and crude oil composition, stable dispersion of asphaltenes in crude oil would be destabilized; hence, asphaltenes precipitate out of the crude oil solution and settle in the form of solid deposits which can impede the production process (Shirani et al. 2012a, b). In upstream operations, precipitation and deposition of asphaltenes in porous medium during natural depletion and EOR processes substantially affect porosity and permeability of the reservoir rock, and consequently results in loss of efficiency of the recovery process (Kord et al. 2012, 2014; Zanganeh et al. 2012; Moradi et al. 2012). Asphaltene precipitation and deposition also arise in production facilities and transportation pipelines which result in considerable increase in cost of production operation (Abu Tarboush and Husein 2012). To alleviate problems associated with asphaltene precipitation and deposition, two fundamental questions are posed: under what conditions asphaltenes will precipitate out of the crude oil? and what is the extent of asphaltene precipitation and deposition under particular circumstances? Several researchers

attempted to identify mechanisms responsible for asphaltenes precipitation and deposition as well as to develop a potent and reliable representative model to quantitatively and qualitatively estimate the extent of asphaltene precipitation and deposition. However, no practical and accurate model is proposed yet to clarify the phase behavior of asphaltene precipitation mainly due to its complex structure and properties (Shirani et al. 2012a, b). The available models for delineation of asphaltenes precipitation are categorized into two distinct classes: models that involve the use of asphaltene properties and the ones that are based on scaling approaches.

Models adopting the asphaltenes properties can be subsumed into two broad categories:

- Crude oil is a real solution with asphaltenes dissolved in it: The major assumption in these models is that asphaltene precipitation is a reversible process (Mansoori 1997).
- Asphaltenes are suspended in crude oil because they are peptized by resins: In these models, asphaltene precipitation is considered an irreversible process (Wu et al. 1998).

The second group of the predictive models is the one which is based on scaling equations. In these models, asphaltene precipitation is predicted without being concerned about the complex nature of the asphaltene agglomerates (Rassamdana et al. 1999).

Intelligence systems have been recently used as mathematical predictive tools to quantitatively formulate the amount of asphaltene precipitation using titration data (Ashoori et al. 2010; Naimi et al. 2014; Ahmadi 2011, 2012; Asoodeh et al. 2014a, b; Abedini et al. 2010; Zahedi et al. 2009; Chamkalani et al. 2013; Gholami et al. 2013; Gholami et al. 2014a, b; Fatahi et al. 2014). Although these predictive models are valuable, the quest for greater accuracy has been always remained an issue. In this paper, ACE was firstly used as a potent model to estimate the amount of asphaltene precipitation based on a series of titration experiments using stock tank oils. This non-parametric statistical regression method enhances the fitting by suggesting the optimal transformations for predictor variables (i.e., titration data including temperature, solvent molar weight, and dilution ratio in our case study) and response variable (i.e., the amount of asphaltene precipitation for the case under study in this paper) (Malallah et al. 2006). Secondly, the SVR method was used as a supervised learning method for predicting the amount of asphaltene precipitation. The SVR method is based on the employment of the structural risk minimization principle; hence, it provides better generalization performance in comparison to the conventional Artificial Neural Networks (ANNs) in which the empirical risk minimization principle is

implemented. In the end, PLCM method was proposed to combine the outputs associated with SVR and ACE models. To assign appropriate weight factors to each of these two models, GA tool was employed as the combiner. The weight factor associated with each individual model demonstrates the contribution of that model in the final quantitative estimation of asphaltene precipitation. The predicted asphaltene precipitation results using these three models were compared with the experimental data to validate the effectiveness and robustness of the proposed models.

Materials and methods

In this section, application of each of the three predictive methods of ACE, SVR, and PLCM in estimating the amount of asphaltene precipitation is described followed by a description on input/output data space used in this study.

Alternating conditional expectation (ACE) algorithm

This algorithm was developed by Breiman and Friedman as an advanced statistical technique for solving the regression problems containing an unidentified relationship between the predictor and the response variable (Breiman and Friedman 1985). This method has found growing application in different industries owing to its remarkable performance including its capability in appropriate demonstration of a non-linear relationship, if it does exist, between variables in regression problems as well as providing improved fit to the experimental data in comparison to the conventional linear models (Eissa and Shokir 2007). This method achieves the maximum coefficient of multiple correlation through approximating the optimal transformations for the dependent and independent variables (Tang and Zhou 2013).

Generally, a linear regression model for p independent variables of X_1, X_2, \dots, X_p and a response variable, Y , is given in the following form (Breiman and Friedman 1985):

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i + \varepsilon, \quad (1)$$

in which β_i , $i = 0 - p$ are the regression coefficients to be estimated, and ε is an error term. In ACE algorithm, the relation between $\theta(Y)$ and $\phi_1(X_1), \dots, \phi_p(X_p)$ is computed in lieu of assessing the correlation between Y and X_1, X_2, \dots, X_p which is conducted in the conventional regression analysis. Based on Eq. (1), the general form of non-parametric ACE algorithm is defined in the following form:

$$\theta(Y) = \alpha + \sum_{i=1}^p \phi(X_i) + \varepsilon, \quad (2)$$

in which $\theta(Y), \phi_1(X_1), \dots, \phi_p(X_p)$ are the arbitrary measurable mean-zero functions of Y, X_1, X_2, \dots, X_p , respectively. The main intention of using ACE algorithm is to seek the optimal transformations, $\phi_i^*(X_i), i = 1, \dots, p$ and $\theta^*(Y)$, which lead to the maximum correlation between the transformed dependent variable and sum of the transformed predicted variables. This is equivalent to minimizing the value of error variance (ε^2). The value associated with the error variance of a linear regression of the transformed dependent variable on the sum of the transformed independent variables under the constraint of $E[\theta^2(Y)] = 1$ is given by the following equation (Breiman and Friedman 1985):

$$\varepsilon^2(\theta, \phi_1, \dots, \phi_p) = E \left(\left[\theta(Y) - \sum_{i=1}^p \phi(X_i) \right]^2 \right) / E\theta^2(Y). \quad (3)$$

ACE algorithm minimizes ε^2 by holding $E[\theta^2(Y)] = 1$, $E\theta(Y) = E\phi_1(X_1) = \dots = E\phi_p(X_p) = 0$ through the use of a series of single-function minimizations involving bivariate conditional expectations. Thus, for a given set functions $\phi_1(X_1), \dots, \phi_p(X_p)$, minimization of ε^2 with respect to $\theta(Y)$ yields the following equation:

$$\theta(Y) = E \left[\sum_{i=1}^p \phi(X_i) | Y \right] / \left\| E \left[\sum_{i=1}^p \phi(X_i) | Y \right] \right\|. \quad (4)$$

On the other hand for a given $\theta(Y)$, minimization of ε^2 with respect to a single function $\phi_j(X_j)$ yields the following equation:

$$\phi_j(X_j) = E \left[\theta(Y) - \sum_{i \neq j}^p \phi_i(X_i) | X_j \right]. \quad (5)$$

Having implemented the iteration process of minimizing ε^2 , the real-valued measurable zero-mean functions of $\phi_i(X_i), i = 1, \dots, p$ and $\theta(Y)$ are determined which are based on determination of the optimal transformation of $\phi_i^*(X_i), i = 1, \dots, p$ and $\theta^*(Y)$. In the transformed space, the response and predictor variables are related to each other according to the following equation:

$$\theta^*(Y) = \sum_{i=1}^p \phi_i^*(X_i) + e^*, \quad (6)$$

in which e^* is the error which is not captured by the use of ACE transformations and is assumed to have a normal distribution with zero mean (Breiman and Friedman 1985).

Support vector regression (SVR) algorithm

SVR is a novel learning methodology based upon statistical learning theory which is applicable for solving the complex non-linear regression estimation problems (Asoodeh et al. 2014a, b, c; Bagheripour et al. 2015). This method was first introduced as a robust tool by Vapnik (2000) and has found drastic applicability in the field of regression estimation owing to its many attractive features and higher generalization ability in comparison to traditional ANNs. The difference between generalization performance of SVR and ANN is attributed to the difference in principle of risk minimization implemented in these two models. In SVR, the structural risk minimization is implemented for minimizing the upper bound of generalization error, while ANN takes the advantage of empirical risk minimization to minimize the prediction error (Asoodeh and Bagheripour 2013).

The main role of the SVR algorithm is to estimate a function which computes the functional dependency between an output variable “y,” defined on “R,” and n-dimensional input variable “x.” The decision function $f(x, w)$ can be approximated using the following equation (Kecman 2005):

$$f(x, w) = w^T x + b, \quad (7)$$

in which w and b denote the weight vector and the bias term, respectively, and w^T denotes the transposition of weight matrices.

For computing $f(x, w)$, it is necessary to calculate the value of unknown parameters of w and b . In the SVR algorithm, the following regularized risk function should be minimized in order to determine w and b :

$$R(C) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l L_e(y_i, f_i(x, w)), \quad (8)$$

in which $C \frac{1}{2} \sum_{i=1}^l L_e(y_i, f_i(x, w))$ is empirical error, $\frac{1}{2} \|w\|^2$ is the measure of function flatness, and l is the total number of sample. The constant $C > 0$ is the penalty parameter which computes the trade-off between the empirical error and the model complexity. ε -insensitive loss function $L_e(y_i, f_i(x, w))$ is used for estimation of empirical error and is determined using the following equation (Kecman 2005):

$$|y - f(x, w)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x, w)| \leq \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{Otherwise} \end{cases}. \quad (9)$$

Minimizing Eq. (8) is equivalent to minimizing the following equation:

$$R = \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^l \xi_i + \sum_{i=1}^l \xi_i^* \right). \quad (10)$$

$$\text{Under the constraints of } \begin{cases} y_i - w^T x_i - b \leq \varepsilon + \xi_i, & i = 1, \dots, l \\ w^T x_i + b - y_i \leq \varepsilon + \xi_i^*, & i = 1, \dots, l \\ \xi_i \geq 0, \xi_i^* \geq 0, & i = 1, \dots, l \end{cases}$$

in which ξ and ξ^* are the positive slack variables that are employed for training data points above and below the regression tube, respectively. Slack variables are determined as follows:

$$\xi = |y - f(x, w)| - \varepsilon \quad \text{for data above an } \varepsilon \text{-tube}, \quad (11)$$

$$\xi^* = |y - f(x, w)| - \varepsilon \quad \text{for data below an } \varepsilon \text{-tube}. \quad (12)$$

For solving Eq. (10), a primal Lagrangian variable is formed first, followed by substituting the Karush–Kuhn–Tucker conditions in primal Lagrangian variable. Once substituted, the problem is formulated in dual space in the form of the following function (Kecman 2005):

$$\begin{aligned} \text{Maximize } L_p(\alpha_i, \alpha_i^*) = & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_i^T x_j \\ & - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i. \end{aligned} \quad (13)$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \\ 0 \leq \alpha_i^* \leq C, \quad i = 1, \dots, l \end{cases}$$

in which $\alpha_i, \alpha_i^* \geq 0$ are called Lagrangian multipliers.

From the above equation, one Lagrange multiplier pair is calculated. By the use of this Lagrange multiplier pair, the optimal desired weight factor of the regression hyperplane is determined as follows:

$$w_o = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i. \quad (14)$$

The training points which satisfy $\alpha_i - \alpha_i^* \neq 0$ are called free support vectors. These points lead to determination of the bias term b . Owing to numerical sensitivity of bias term computation, it is preferable to calculate the bias term by averaging over all the free support vectors. Hence, the best hyperplane regression estimation model becomes (Kecman 2005):

$$f(x, w) = w_o^T x + b = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i^T x + b. \quad (15)$$

A generalization of the SVR method to non-linear regression is performed through kernel function. Kernel function maps the input vector $x \in R^n$ into vector $\phi(x)$ of high dimension feature space F and then solves a linear regression problem in this feature space. Several kernel functions are employed in SVR namely linear, polynomial, radial basis function (RBF), and sigmoid. By employment of the kernel function G and the Lagrangian multiplier pairs of α_i, α_i^* , the best non-linear regression function is computed as follows:

$$f(x, w) = G(\alpha - \alpha^*) + b. \quad (16)$$

For determining the bias term b , it is recommended to average over all the free support vectors (Kecman 2005).

Committee machine (CM) algorithm

This method has been recognized as a promising approach in solving estimation problems owing to its better performance in comparison to its individual elements that are executed for prediction target. The essence of the CM algorithm is that it has a parallel framework which produces a final output by combining the results associated with individual models (Asoodeh and Bagheripour 2012a, b; Gholami et al. 2014a, b). Every predictive algorithm such as ANN, empirical correlation, and other modeling algorithms can construct the element of CM (Chen and Lin 2006). By combining the output of individual experts, CM reaps the benefits of each expert. Weight factor of each model indicates its contribution in the final prediction. Optimal contribution of each algorithm in the overall prediction is determined through Genetic Algorithm, GA (Asoodeh, 2013). The schematic diagram of CM is demonstrated in Fig. 1 (Haykin 1999).

Genetic algorithm (GA)

The GA procedure was first introduced by John Holland in 1975 and is a robust optimization technique based on principle of natural selection and evolution (Holland 1975). This method has gained prominence as an effective searching technique owing to its high capability of solving non-linear and discontinuities problems in widespread fields of research. The simple GA cycle comprises of three stages of selection, genetic operation, and replacement. Optimization is started with initial population of random chromosomes. Computation of each chromosome fitness value is fulfilled by objective function(s). The best parent solutions are reproduced and the new population is generated through generic operation. Generic operation employs two processes to generate children namely as crossover and mutation. Crossover permits the exchange of information among individuals in the population and provides the innovative capability of GA. Mutation ensures desirable diversity. The children solutions are evaluated and the whole cycle repeats until desirable termination criterion is attained (Kadkhodaie-Ilkhchi et al. 2009; Ghiassi-Freez et al. 2012). A general flowchart of the GA procedure is shown in Fig. 2.

Input/output data space

The asphaltene precipitation data used to construct the predictive models in this study were borrowed from the open literature (Hu and Guo 2001). The authors added normal alkanes with different carbon numbers to stock tank oil at different dilution ratios and temperatures in the form of a series of titration tests. Asphaltenes precipitated in these titration tests were then separated from the solution

Fig. 1 A schematic diagram of CM (Haykin 1999)

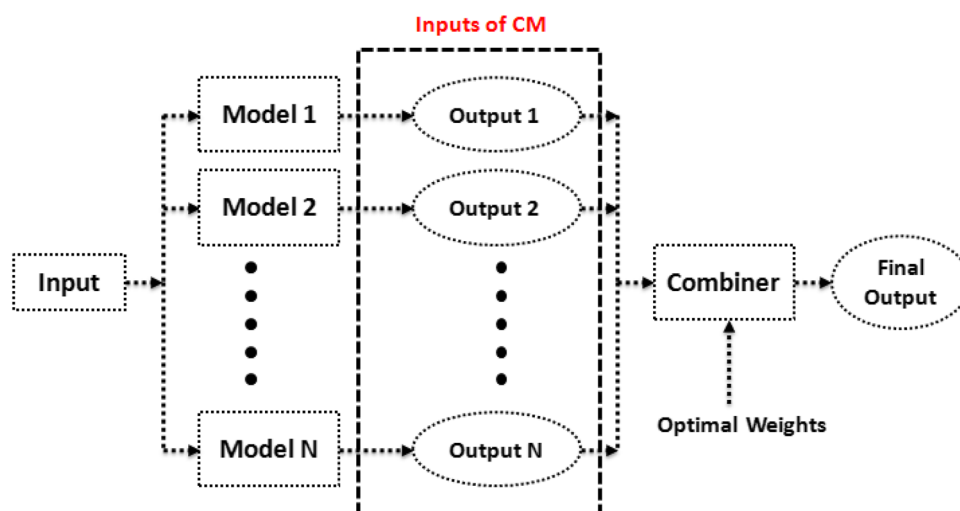
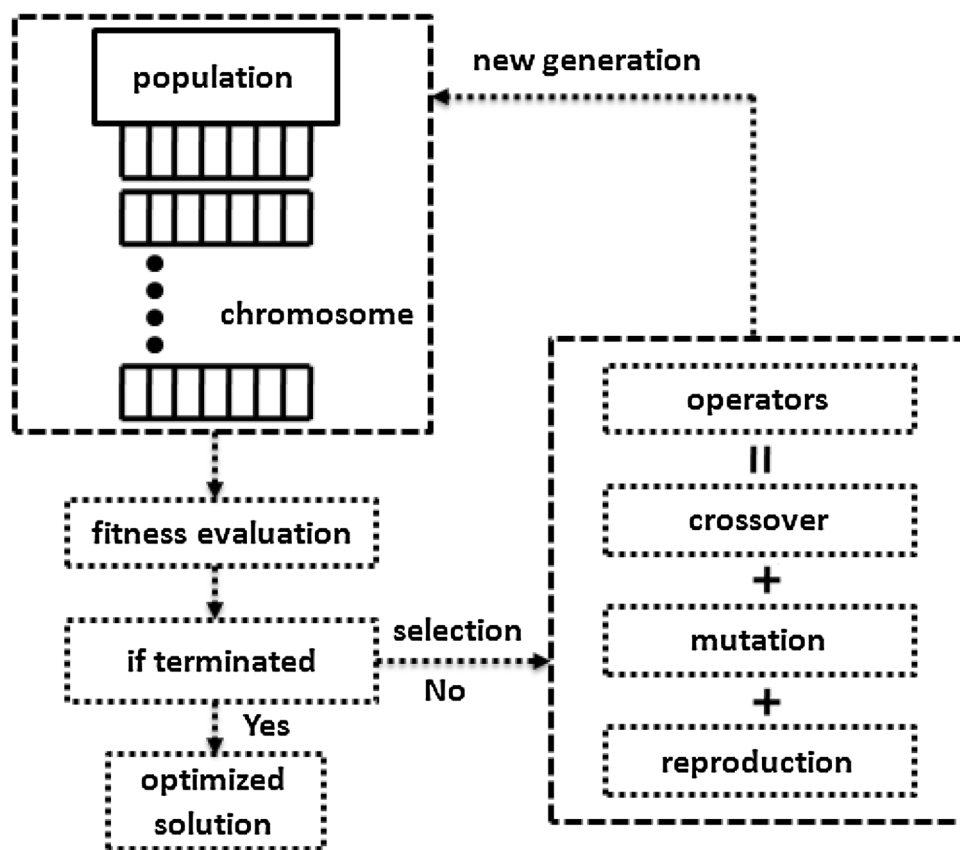


Fig. 2 A general flowchart of GA (Chen and Lin 2006)



using Whatman filter paper No. 42. The composition of oil used in this study is provided in Table 1. The amount of asphaltene precipitation was then modeled using a scaling equation. In this modeling approach, the input variables consist of dilution ratio “ R_v ,” temperature “ T ,” and molecular weight of solvent “ M_w ,” while the amount of asphaltene precipitation “ W_i ” was regarded as the target value. Each of these input variables had qualitative effects on the amount of the target value according to the sets of data used in this study. For instance, increasing the dilution ratio increased the amount of asphaltene precipitation until the entire asphaltene content of the crude oil sample was precipitated out of the solution. By increasing the molecular weight of the solvent, the amount of asphaltene precipitation decreased. Temperature has a dual impact on the amount of asphaltene precipitation. While many researchers contended that an increase in temperature would increase the amount of asphaltene precipitation, some others believe the reverse trend would be the case. According to Hu and Guo’s study, increasing the temperature decreased the amount of asphaltene precipitation. The data collected from the literature was divided into two classes of training data and testing data points. 70 % of the data points (i.e., 124 data points in total) were employed to

construct the model whereas the remainder 30 % (i.e., 52 data points in total) was used to test the model. The statistics of the data used in this study is provided in Table 2.

In order to check the validity and accuracy of the asphaltene precipitation data predicted using PLCM model with ACE and SVR as the constituents, two statistical parameters were defined namely as Correlation Coefficient (R^2) and Mean Square Error (MSE). Correlation Coefficient (R^2) illustrates how good the model corresponds the experimental data, hence represents a measure of the utility of the model. The closer the value of R^2 to 1, the better the model fits the data. This parameter is mathematically represented by

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{meas}} - Y_{\text{pred}})^2}{\sum_{i=1}^n (\bar{Y} - Y_{\text{pred}})^2}, \quad (17)$$

in which Y_{meas} , Y_{pred} , \bar{Y} , and n are measured variables, estimated variables, mean of measured variables, and number of data points, respectively.

Mean Square Error (MSE) computes the data dispersion around zero deviation and is defined as

Table 1 Compositional characteristics and properties of the degassed Caoqiao crude oil and separator gas (Hu and Guo 2001)

Component	Degassed oil	Separator gas
CO ₂	0	2.96
N ₂	0	1.18
C ₁	0	89.37
C ₂	0	3.34
C ₃	0	2.10
i-C ₄	0	0.32
n-C ₄	0	0.26
i-C ₅	0.16	0.22
n-C ₅	0.58	0.15
n-C ₆	1.2	0.12
C ₇₊	98.06	N/A
C ₁₁₊	87.16	N/A
C ₇₊ molecular weight (g/mol)	503.6	
C ₇₊ density (at 293 K)	0.9526	
Reservoir conditions		
Reservoir temperature (K)	343	
Bubble point pressure at 343 K (MPa)	9.8	
Gas oil ratio (m ³ /m ³)	30.2	
SARA analysis of the stock tank oil		
Saturates (wt%)	38.0	
Aromatics (wt%)	47.6	
n-C ₅ Asphaltene (wt%)	7.26	
Resins (wt%)	18.6	

Table 2 Statistical description of the dataset used for developing the predictive models

Parameter	Min	Max	Mean
Dilution ratio (mL/g) (R_v)	2.3	24.3	12.08
Temperature (K) (T)	293	338	312.43
Molecular weight of solvent (g/mol) (M_w)	72.15	170.33	116.14
Amount of asphaltene precipitation (wt%) (W_t)	0.12	7.06	2.97

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_{i\text{meas}} - Y_{i\text{pred}})^2. \quad (18)$$

Results and discussion of results

ACE model predictions

The ACE procedure was first employed to construct the model. To achieve the best correlation between the predictor variables and the response value, the ACE procedure has the capability of using the independent and dependent

variables not only in their actual space but also in their logarithmic space. After testing all possible combinations of the predictor and response variables in either logarithmic or actual space, the optimal combination is computed as follows:

$$W_t = f(T, R_v, M_w). \quad (19)$$

The optimal transformations for temperature, dilution ratio, solvent molecular weight, and amount of asphaltene precipitation, calculated by the use of ACE algorithm, are presented in Fig. 3a–d. These figures helped to extract the formulation between the values of certain parameters and optimal transformation associated with them. The optimal transformation of W_t versus summation of the optimal transformations of T , R_v , and M_w is presented in Fig. 4. The deviation of some data points from the straight line correlation demonstrates the error which could not be captured by ACE transformation method. Furthermore, the comparison cross-plot of the amount of asphaltene precipitation which is predicted using the ACE procedure versus the corresponding set of experimental data is presented in Fig. 5. The acceptable agreement between the predicted and actual measurements is observed.

To develop a computational model for predicting the amount of asphaltene precipitation, it seems crucial to seek a correlation between the optimal data transforms and the actual data values. For this purpose, the optimal data transforms were related to the original data through the use of curve fitting. The resulting equations for temperature, dilution ratio, and molecular weight of solvent with respect to their corresponding coefficients are

$$\phi^*(R_v) = a_1 \text{EXP}(a_2 \times R_v) + a_3 \text{EXP}(a_4 \times R_v), \quad (20)$$

in which $a_1 = 0.28350000$, $a_2 = 0.05531000$, $a_3 = -4.858$, and $a_4 = -0.29660000$,

$$\phi^*(M_w) = a_3(M_w)^2 + a_2(M_w) + a_1, \quad (21)$$

in which $a_3 = 0.00031400$, $a_2 = -0.22320000$, and $a_1 = 39.02000000$,

$$\phi^*(T) = a_3(T)^2 + a_2(T) + a_1, \quad (22)$$

in which $a_3 = 0.00031400$, $a_2 = -0.22320000$, and $a_1 = 39.02000000$.

The linear regression between the above independent variables and the dependent variable of the amount of asphaltene precipitation is given as follows

$$\theta^*(W_t) = \phi^*(T) + \phi^*(R_v) + \phi^*(M_w). \quad (23)$$

The numerical value of the amount of asphaltene precipitation is determined by employing the following polynomial correlation:

$$W_t = a_3(\theta^*(W_t))^2 + a_2(\theta^*(W_t)) + a_1, \quad (24)$$

Fig. 3 Optimal transformation of **a** dilution ratio, **b** molecular weight of solvent, **c** temperature, and **d** amount of asphaltene precipitation computed by ACE

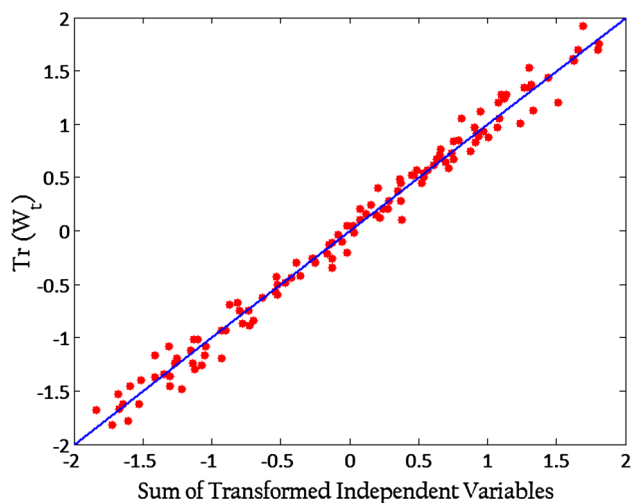
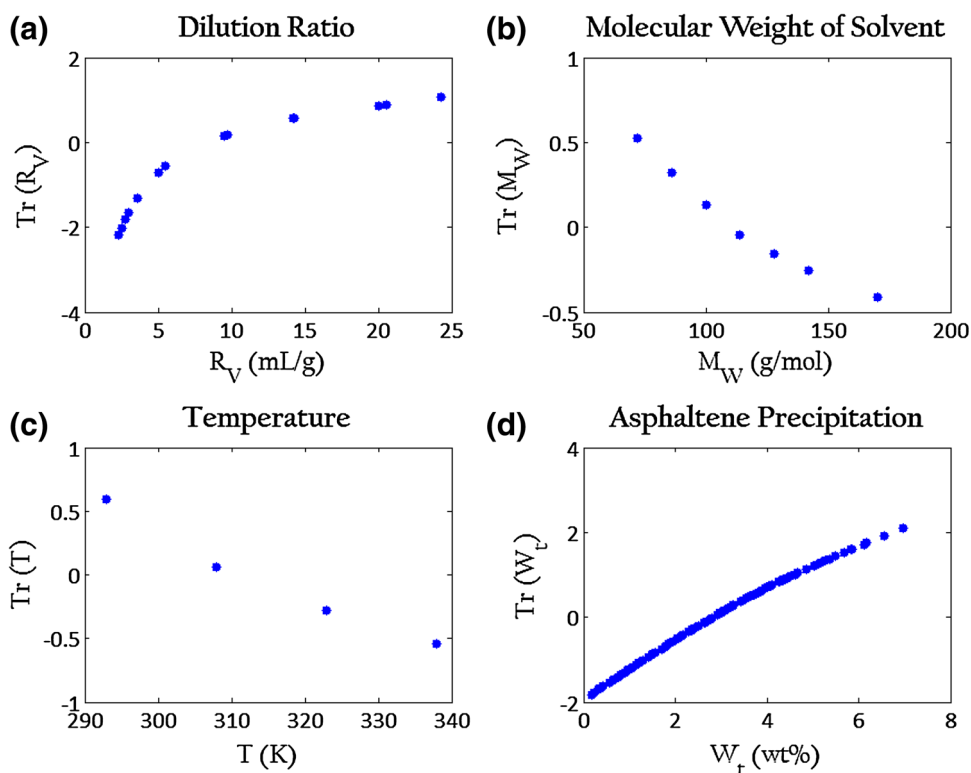


Fig. 4 Optimal transformation of W_t versus the sum of the optimal transformations of R_v , M_w , and T

in which $a_3 = 0.13230000$, $a_2 = 1.65900000$, and $a_1 = 2.81900000$.

Table 3 illustrates the R^2 and MSE values for the ACE model for the training and testing stages. Moreover, a comparison between the measured and predicted amounts of asphaltene precipitation versus different sample

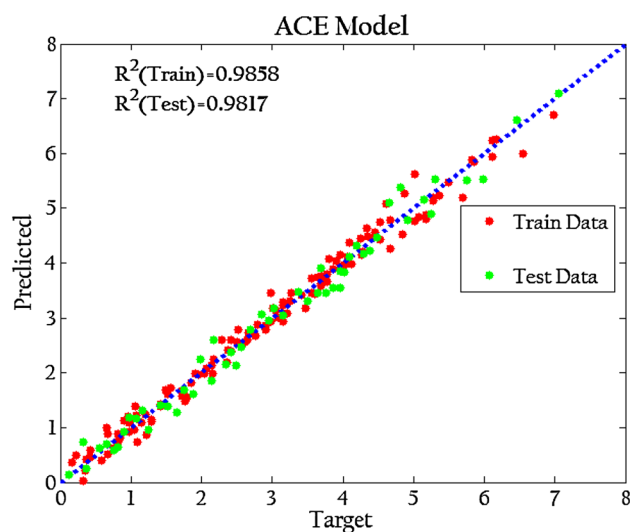


Fig. 5 Cross-plot illustrating the correlation coefficient between measured asphaltene precipitation values and those predicted by the ACE procedure

numbers is demonstrated in Fig. 6. According to the data presented in Figs. 5, 6 and also in Table 3, ACE model has a good performance in terms of predicting the amount of asphaltene precipitation.

Table 3 Comparison between PLCM model and the individual ACE and SVR models in term of the statistical parameters for the training and testing stages

Model	Allocation	R^2	MSE
ACE	Training	0.9858	0.0390
ACE	Testing	0.9817	0.0550
SVR	Training	0.9929	0.0197
SVR	Testing	0.9893	0.0321
PLCM	Training	0.9960	0.0111
PLCM	Testing	0.9917	0.0248

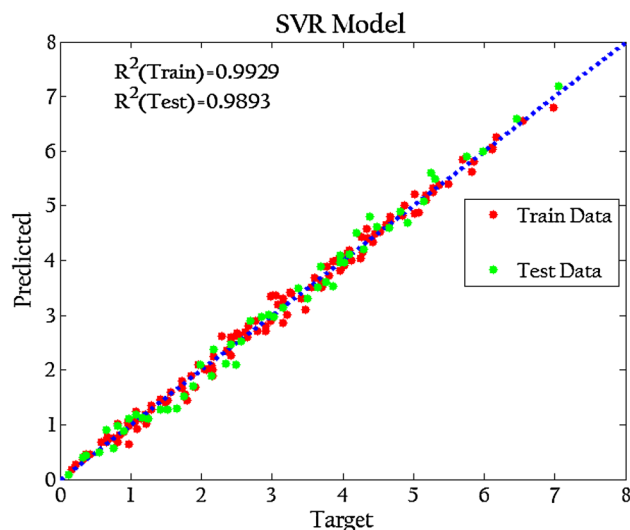
SVR model predictions

A detailed study about formulating the values of asphaltene precipitation associated with the titration data through the use of SVR was performed by Naimi et al. (2014). The same analogy was used in this section using our selected set of experimental data borrowed from Hu and Guo (2001). An epsilon SVR model was employed in this section to predict the amount of titration-based asphaltene precipitation based on a series of experimental data. Radial basis function (RBF) was chosen as the kernel function owing to its simplicity and greater performance. Optimum range for RBF performance is $[-1 \ 1]$ prior to applying the constructed model. Grid search and pattern search are some surveying techniques for computing the optimal values of parameters involved in RBF namely as C, Gamma, and Epsilon. The search range and extracted optimal values for these involved parameters are reported in Table 4.

To evaluate the performance of the SVR model, titration data from Hu and Guo (2001) were used as inputs for the developed model and the amount of asphaltene precipitation was predicted. The cross-correlation between the

Table 4 Properties of optimal SVR model

Parameters	C	Gamma	Epsilon
Search ranges	0.1–500,000	0.000001–20	0.0001–100
Optimum value	67855.53522	0.25516	0.01265

**Fig. 7** Cross-plot illustrating the correlation coefficient between measured asphaltene precipitation values and those predicted by the SVR model

measured and SVR-predicted amount of asphaltene precipitation is illustrated in Fig. 7. The statistical parameters of R^2 and MSE for the SVR model predictions are also shown in Table 3. Significantly high value of correlation coefficient as well as low values associated with MSE confirms the success of SVR modeling in estimating the

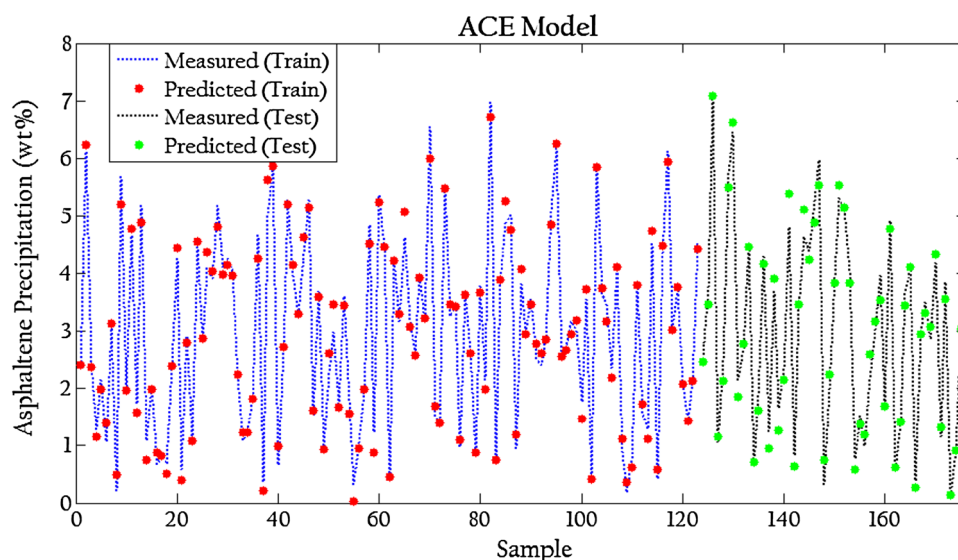
Fig. 6 A comparison between the measured amounts of asphaltene precipitation and the predicted values using the ACE model

Fig. 8 A comparison between the measured amounts of asphaltene precipitation and the predicted values using the SVR model

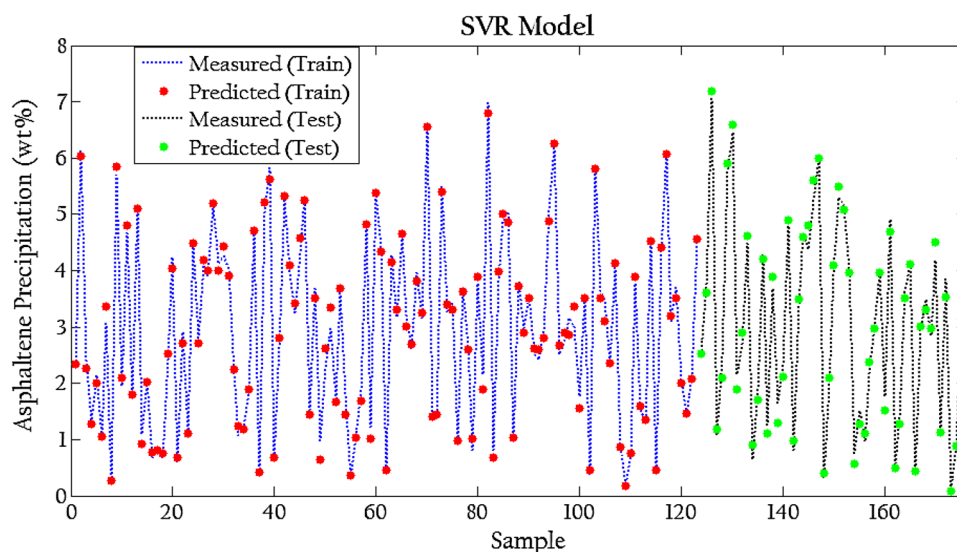
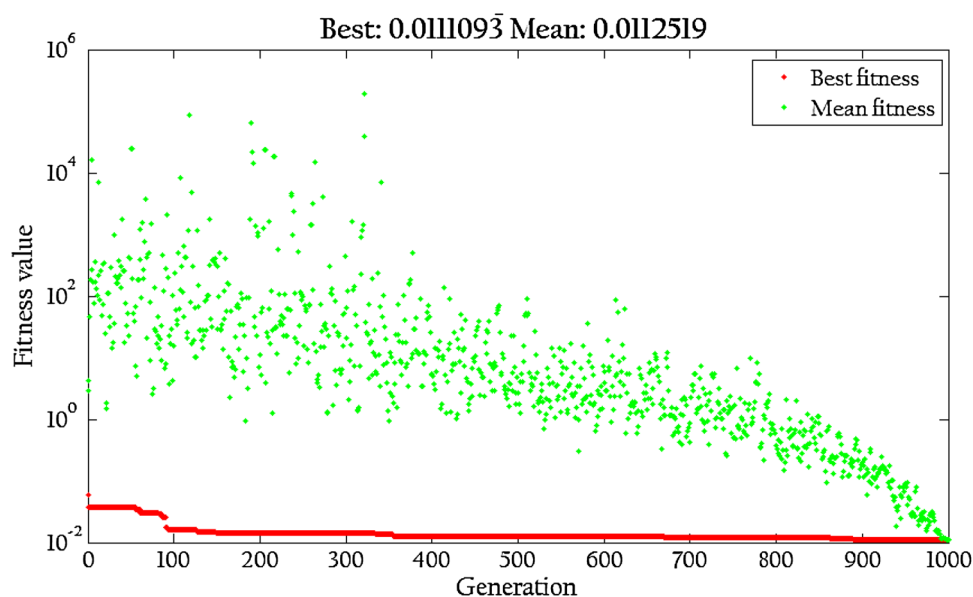


Fig. 9 The best and mean predicted fitted values associated with the amount of asphaltene precipitation after 1000 generations



amount of asphaltene precipitation. A comparison between the measured and predicted values of asphaltene precipitation for different sample numbers is also shown in Fig. 8. This figure shows that SVR is a potent model to predict the amount of asphaltene precipitation.

Power-law committee machine (PLCM) model predictions

A PLCM model was developed in this study by adopting the concept of “divide and conquer” to combine the outputs of ACE and SVR models. Different studies showed that breaking down the data space into subspaces and solving each separately would lead to increasing accuracy

of final prediction (Asoodeh and Bagheripour 2012a, b). To construct the proposed PLCM model, the following equation was introduced to GA:

$$\text{MSE}_{\text{PLCM}} = \frac{1}{k} \sum_{i=1}^k (w_1 \times \text{ACE}_i^{w_2} + w_3 \times \text{SVR}_i^{w_4} - T_i)^2, \quad (25)$$

in which w_i ($i = 1, \dots, 4$) refer to weight factors assigned to each particular output of the predictive models, k denotes the total number of data, and T_i refers to the corresponding target value.

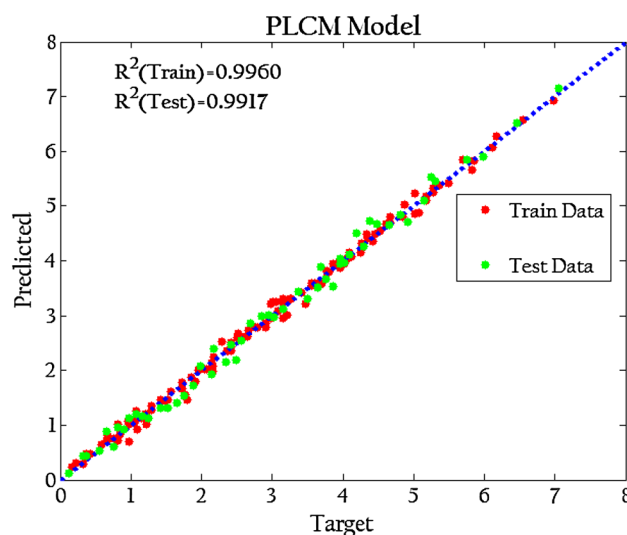
The GA which is implanted in the body of the CM model is able to extract optimum values of the aforementioned weight factors in such a way that the CM model would cover to the lowest MSE associated with the

Table 5 Weights assigned to each model's contribution to optimize the final prediction

Weight	W_1	W_2	W_3	W_4
Value	0.175	0.578	0.843	1.05

prediction. GA initiates with random population of 50 potential chromosomes (i.e., solutions) and evolves toward the best chromosome through the use of genetic operations. Genetic operators, including elite preservation and cross-over were regulated to 2 and 0.76, respectively. In other words, 20 % of the succeeding generation is composed of mutation and inversion operations. The execution of the GA procedure for the proposed PLCM model to predict titration-based asphaltene precipitation is shown in Fig. 9.

After running the GA procedure, the optimal weight factors associated with PLCM model were extracted. Weights which were assigned to each model in Eq. 25 are shown in Table 5. These values demonstrate the optimal contribution of each model in final predictions. By applying these weight factors to the outputs of the individual models of ACE and SVR, the amount of asphaltene precipitation for new titration datasets was predicted and then was compared against the experimental data. A diagram of the proposed PLCM model for the purpose of predicting

**Fig. 11** Cross-plot illustrating the correlation coefficient between measured asphaltene precipitation values and those predicted by the PLCM model

the amount of asphaltene precipitation is shown in Fig. 10. To assess the reliability of the constructed PLCM model, asphaltene precipitation testing datasets were used as input feed data for the PLCM model. The cross-plot composed of the predicted values of asphaltene precipitation versus the

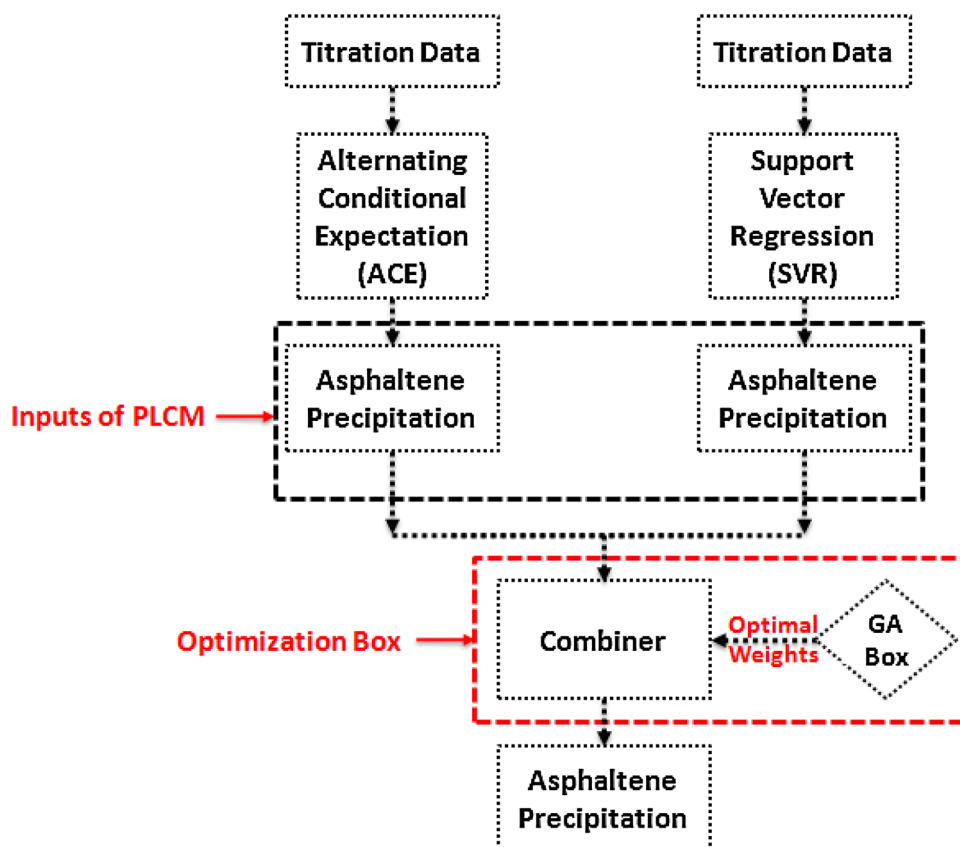
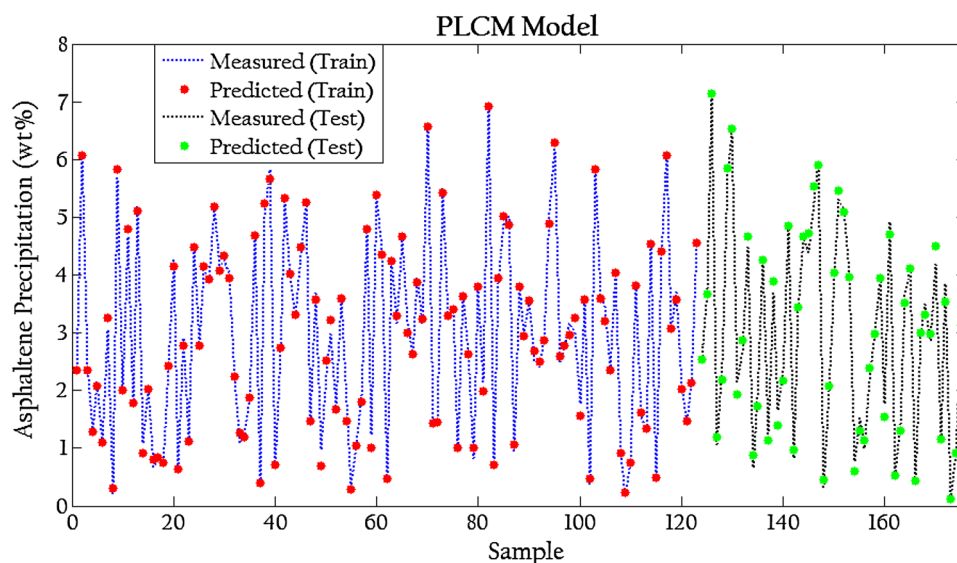
Fig. 10 Schematic diagram of the PLCM model designed in this study

Fig. 12 A comparison between the measured amounts of asphaltene precipitation and the predicted values using the PLCM model



measured amounts from the titration experiments is shown in Fig. 11.

The comparison between the statistical parameters representing the performance of the PLCM model as opposed to the individual models of ACE and SVR is provided in Table 3. The trends presented in Fig. 11 as well as the statistical indicators provided in Table 3 prove that the PLCM predicted results are more reliable compared to the other individual models. In other words, in situations where there are multiple options to solve a problem, it is possible to construct a PLCM model by spending little additional computation effort to the expense of having enhanced accuracy in final predictions. In addition, Fig. 12 demonstrates a comparison between the measured and predicted amounts of asphaltene precipitation versus different data samples. From this Figure, it is concluded that PLCM model is a powerful tool for predicting the amount of asphaltene precipitation in titration experiments.

Conclusions

Three smart techniques were used in this study to predict the amount of asphaltene precipitated during titration experiments. It was concluded that PLCM model is capable of predicting the amount of asphaltene precipitation with higher accuracy considering the statistical analysis performed on each prediction. The results predicted using the PLCM model were compared against those obtained using the other two methods of ACE and SVR which proved the superior performance of the PLCM model compared to the other two predicting procedures.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abedini A, Ashoori S, Saki Y (2010) Application of Neural Network Model for Prediction of Asphaltene Precipitation. SPE-132760-MS
- Abu Tarboush BJ, Husein MM (2012) Adsorption of asphaltenes from heavy oil onto in situ prepared NiO nanoparticles. J Colloid Interface Sci 378:64–69
- Ahmadi MA (2011) Prediction of asphaltene precipitation using artificial neural network optimized by imperialist competitive algorithm. J Pet Explor Prod Technol 1:99–106
- Ahmadi MA (2012) Neural network based unified particle swarm optimization for prediction of asphaltene precipitation. Fluid Phase Equilib 314:46–51
- Ahmadi MA, Shadizadeh SR (2012) New approach for prediction of asphaltene precipitation due to natural depletion by using evolutionary algorithm concept. Fuel 102:716–723
- Ashoori S, Abedini A, Abedini R, Qorbani Nasheghi K (2010) Comparison of scaling equation with neural network model for prediction of asphaltene precipitation. J Pet Sci Eng 72:186–194
- Asoodeh M (2013) Prediction of Poisson's Ratio from Conventional Well Log Data: a Committee Machine with Intelligent Systems Approach. Energy Sources Part A 35:962–975
- Asoodeh M, Bagheripour P (2012a) Prediction of compressional, shear and stoneley wave velocities from conventional well log data using a committee machine with intelligent systems. Rock Mech Rock Eng 45(1):45–63
- Asoodeh M, Bagheripour P (2012b) Estimation of bubble point pressure from PVT data using a power-law committee with intelligent systems. J Pet Sci Eng 90–91:1–11
- Asoodeh M, Bagheripour P (2013) Fuzzy classifier based support vector regression framework for Poisson ratio determination. J Appl Geophys 96:7–10

- Asoodeh M, Gholami A, Bagheripour P (2014a) Renovating scaling equation through hybrid genetic algorithm-pattern search tool for asphaltene precipitation modeling. *J Dispers Sci Technol* 35(4):607–611
- Asoodeh M, Gholami A, Bagheripour P (2014b) Asphaltene precipitation of titration data modeling through committee machine with stochastically optimized fuzzy logic and optimized neural network. *Fluid Phase Equilib* 364:67–74
- Asoodeh M, Gholami A, Bagheripour P (2014c) Oil-CO₂ MMP Determination in Competition of Neural Network, Support Vector Regression, and Committee Machine. *J Dispers Sci Technol* 35(4):564–571
- Bagheripour P, Gholami A, Asoodeh M, Vaezzadeh-Asadi M (2015) Support vector regression based determination of shear wave velocity. *J Pet Sci Eng* 125:95–99
- Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation. *J Am Stat Assoc* 80(391):580–598
- Chamkalani A, Amani M, Kiani MA, Chamkalani R (2013) Assessment of asphaltene deposition due to titration technique. *Fluid Phase Equilib* 339:72–80
- Chen CH, Lin ZS (2006) A committee machine with empirical formulas for permeability prediction. *Comput Geosci* 32(4):485–496
- Eissa M, Shokir EM (2007) CO₂–oil minimum miscibility pressure model for impure and pure CO₂ streams. *J Pet Sci Eng* 58:173–185
- Fatahi H, Gholami A, Amiribakhtiar S, Moradi S (2014) Estimation of asphaltene precipitation from titration data: a hybrid support vector regression with harmony search. *Neural Comput Appl*. doi:10.1007/s00521-014-1766-y
- Ghiasi-Freez J, Kadhodaie-Ilkhchi A, Ziaii M (2012) Improving the accuracy of flow units prediction through two committee machine models: an example from the South Pars Gas Field, Persian Gulf Basin, Iran. *Comput Geosci* 46:10–23
- Gholami A, Moradi S, Dabir B (2013) A Power Law Committee Scaling Equation for Quantitative Estimation of Asphaltene Precipitation. *Int J Sci Emerging Technol* 5(6):275–283
- Gholami A, Moradi S, Asoodeh M, Bagheripour P, Vaezzadeh-Asadi M (2014a) Asphaltene precipitation modeling through ACE reaping of scaling equations. *Sci China Chem* 57(12):1774–1780
- Gholami A, Asoodeh M, Bagheripour P (2014b) How committee machine with SVR and ACE estimates bubble point pressure of crudes. *Fluid Phase Equilib* 382:139–149
- Haykin S (1999) *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice-Hall
- Holland JH (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor
- Hu YF, Guo TM (2001) Effect of temperature and molecular weight of n-alkane precipitants on asphaltene precipitation. *Fluid Phase Equilib* 192(1–2):13–25
- Kadkhodaie-Ilkhchi A, Rahimpour-Bonab H, Rezaee M (2009) A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: an example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran. *Comput Geosci* 35(3):459–474
- Kecman V (2005) Support vector machines—an introduction. In: Wang L (ed) *Support Vector Machines: Theory and Applications*, Chap. 1. Springer, Berlin, Heidelberg, pp 1–47
- Kord S, Miri R, Ayatollahi S, Escrochi M (2012) Asphaltene deposition in carbonate rocks: experimental investigation and numerical simulation. *Energy Fuels* 26(10):6186–6199
- Kord S, Mohammadzadeh O, Miri R, Soulgani BS (2014) Further investigation into the mechanisms of asphaltene deposition and permeability impairment in porous media using a modified analytical model. *Fuel* 117PA:259–268
- Kurup A, Wang J, Subramani HJ, Buckley J, Creek JL, Chapman WG (2012) Revisiting Asphaltene Deposition Tool (ADEPT): field application. *Energy Fuels* 26:5702–5710
- Malallah A, Ghorbi R, Algharib M (2006) Accurate estimation of the World Crude Oil PVT Properties using graphical alternating conditional expectation. *Energy Fuels* 20:688–698
- Mansoori GA (1997) Modeling of asphaltene and other heavy organic depositions. *J Pet Sci Eng* 17:101–111
- Mohammadi AM, Eslamimanesh A, Gharagheizi F, Richon D (2012) A novel method for evaluation of asphaltene precipitation titration data. *Chem Eng Sci* 78:181–185
- Moradi S, Dabir B, Rashtchian D, Mahmoudi B (2012) Effect of Miscible Nitrogen Injection on Instability, Particle Size Distribution, and Fractal Structure of Asphaltene Aggregates. *J Dispers Sci Technol* 33:763–770
- Naimi SR, Gholami A, Asoodeh M (2014) Prediction of Crude Oil Asphaltene Precipitation Using Support Vector Regression. *J Dispers Sci Technol* 35(4):518–523
- Rassamdana H, Farhani M, Dabir B, Mozaffarian M, Sahimi M (1999) Asphalt Flocculation and Deposition. V. Phase behavior in miscible and immiscible injections, energy and fuels 13:176–187
- Rasuli Nokandeh N, Khishvand M, Naseri A (2012) An artificial neural network approach to predict asphaltene deposition test result. *Fluid Phase Equilib* 329:32–41
- Shirani B, Nikazar M, Naseri A, Mousavi-Dehghani SA (2012a) Modeling of asphaltene precipitation utilizing Association Equation of State. *Fuel* 93:59–66
- Shirani B, Nikazar M, Mousavi-Dehghani SA (2012b) Prediction of asphaltene phase behavior in live oil with CPA equation of state. *Fuel* 97:89–96
- Tang X, Zhou J (2013) Nonlinear relationship between the real exchange rate and economic fundamentals: evidence from China and Korea. *J Int Money Finance* 32:304–323
- Vapnik V (2000) *The Nature of Statistical Learning Theory*, 2nd edn. Springer, New York
- Wu J, Prausnitz JM, Firoozabadi A (1998) Molecular-thermodynamic framework for asphaltene-oil equilibria. *AIChE* 44(5):1188–1199
- Zahedi G, Fazlali AR, Hosseini SM, Pazuki GR, Sheikhattar L (2009) Prediction of asphaltene precipitation in crude oil. *J Petrol Sci Eng* 68:218–222
- Zanganeh P, Ayatollahi S, Alamdari A, Zolghadr A, Dashti H, Kord S (2012) Asphaltene deposition during CO₂ injection and pressure depletion: a visual study. *Energy Fuels* 26(2):1412–1419